

IBM accélère en matière d'IA d'entreprise pour les clients avec de nouvelles fonctionnalités sur l'IBM Z



Par [Elpida Tzortzatos Meeta Vouk](#), le 28 septembre 2023 : Aujourd'hui, nous sommes ravis de dévoiler une nouvelle suite d'offres d'IA pour l'IBM Z, conçue pour aider les clients à améliorer leurs résultats commerciaux en accélérant la mise en œuvre de l'IA d'entreprise sur l'IBM Z dans une grande variété de cas d'usage et d'industries. Nous apportons [l'intelligence artificielle \(IA\)](#) à des cas d'usage émergents que nos clients (tels que l'assureur suisse La Mobilière) ont commencé à explorer, comme l'amélioration de la précision des recommandations en matière de politique d'assurance, l'amélioration de la précision et de la rapidité des contrôles anti-blanchiment d'argent et la réduction de l'exposition au risque pour les sociétés de services financiers.

Les nouvelles offres sont conçues pour aider à accélérer l'adoption de frameworks et outils d'IA répandus. Nous annonçons également une version améliorée de Machine Learning pour z/OS, ainsi que des améliorations avancées en matière d'intelligence et d'exploitation avec le [dernier système d'exploitation IBM z/OS](#).

Nos [offres riches en matière d'IA](#), associées à une inférence à faible latence qui exploite l'accélérateur d'IA sur la puce IBM Telum, sont conçues pour aider les clients à intégrer des applications d'IA dans leurs applications métier les plus critiques exécutées sur l'[IBM Z](#) et à respecter leurs accords de niveau de service (SLA).

De nouvelles fonctionnalités

- **[Série d'outils d'IA pour IBM Z et LinuxONE](#)** : cette série d'outils, qui devrait être disponible au quatrième trimestre, devrait prendre en charge les frameworks open-source standard du secteur, tels que IBM Z Accelerated for TensorFlow, IBM Z Accelerated for TensorFlow Serving, IBM Z Accelerated for Snap ML, afin que les entreprises puissent commencer à mettre en œuvre une IA digne de confiance.
- **[Série d'outils Python d'IA pour IBM z/OS](#)** : cette [série d'outils améliorée](#) est une bibliothèque de logiciels Python open-source pour prendre en charge les applications d'IA et de [machine learning](#) qui adhèrent aux pratiques d'[IBM en matière de sécurité et de confidentialité dès la conception](#). Les clients

peuvent désormais utiliser des packages éligibles au zIIP avec Python AI Toolkit for IBM z/OS et IBM Open Enterprise SDK for Python 3.11 pour intégrer l'IA dans leurs applications.

Pour aider les data scientists, les développeurs et les équipes informatiques à implémenter l'IA ensemble, ces séries d'outils pour le Z sont conçues pour aider les clients à connecter les données et les applications des mainframes à des frameworks et des packages open-source. Ces outils comprennent des frameworks et des bibliothèques optimisés et pris en charge pour l'IBM Z, et ils sont conçus pour permettre aux développeurs de commencer à mettre en œuvre des capacités d'IA fiables sur z/OS. Ces outils se basent sur les fondements d'[IBM watsonx.ai](https://www.ibm.com/watsonx/ai), plateforme qui permet aux data scientists et aux développeurs de créer, d'exécuter et de gérer des modèles de machine learning. Ces séries d'outils évolueront au fil du temps pour devenir une partie intégrante de la plateforme [IBM watsonx](https://www.ibm.com/watsonx/).

- **[Machine Learning pour IBM z/OS](#) - Éditions Enterprise et Core** : la version améliorée de Machine Learning for IBM z/OS est la plateforme phare d'IBM pour le cycle de vie complet, conçue pour aider les organisations à créer, déployer, gérer et rendre opérationnels les modèles de machine learning et de deep learning sur z/OS. Créée pour les développeurs et les data scientists, cette plateforme est une extension de watsonx.ai pour z/OS et est conçue pour accélérer le développement, le déploiement et le contrôle des modèles de machine learning. Selon nous, les organisations doivent être claires sur la façon dont leurs modèles d'IA sont entraînés, sur les données utilisées dans cette entraînement et sur ce qui entre dans les recommandations d'un modèle d'IA. Aujourd'hui, les clients peuvent mettre en œuvre des capacités d'IA fiables sur l'IBM Z grâce aux capacités d'[IBM Cloud Pak for Data](#), ce qui permet de s'assurer que les modèles et les workflows sont transparents et explicables. Dans les mois à venir, IBM prévoit de déployer ces capacités de manière native pour les applications sur IBM z/OS.
- **[Cloud Pak for Data sur IBM Z](#)** : il s'agit d'un outil d'IA automatique amélioré et puissant au sein de Cloud Pak for Data 4.7 pour automatiser le processus de création de modèles d'apprentissage automatique. Il permet aux utilisateurs de télécharger leurs données, de choisir le type de problème, de spécifier les contraintes et d'exécuter une série d'expériences automatisées qui génèrent rapidement et facilement une gamme de pipelines très performants.
- **[IBM z/OS 3.1 infusé d'IA](#)** : disponible sur le marché le 29 septembre 2023, IBM z/OS 3.1 représente une nouvelle ère dans l'intelligence du système d'exploitation. Grâce aux nouveaux services d'IA pour les systèmes pour IBM z/OS, le système est conçu pour apprendre et prédire comment optimiser les processus informatiques, simplifier la gestion, améliorer les performances et réduire les besoins en compétences particulières.

Comment La Mobilière façonne le futur de l'assurance grâce à l'IA

Nous nous sommes engagés à optimiser la plateforme IBM Z pour l'IA à partir du silicium, en commençant par le processeur IBM Telum. L'[IBM z16](#) offre un accélérateur intégré sur la puce pour une inférence à faible latence,

soutenue par la résilience et la sécurité de l'IBM Z. En se concentrant sur la réduction de la latence pendant l'inférence, les clients peuvent réduire le temps nécessaire pour recevoir des informations et des recommandations de la part des modèles d'IA.

*« D'après nos [recherches](#), nous estimons qu'environ 70 % des transactions mondiales, en valeur, s'effectuent sur l'IBM Z », a déclaré **Neil Katkov, PhD, Director chez Celent**. « L'application du réglage fin et de l'inférence aux données critiques à cette échelle constitue une remarquable innovation pour les cas d'usage dans les opérations commerciales et informatiques. »*

L'analyse des transactions à grande échelle pourrait permettre aux compagnies d'assurance de formuler plus rapidement des recommandations en matière d'offres d'assurance. Par exemple, La Mobilière s'est appuyée sur l'IBM Z pour traiter d'importants volumes de transactions et fournir un environnement sécurisé pour ses données les plus sensibles. L'utilisation de SQL Data Insights sur l'IBM z16 lui a permis de découvrir des modèles dans ses données critiques afin d'améliorer ses prévisions d'offres d'assurance et de réduire les coûts.

*« Dans le secteur de l'assurance, les clients recherchent le service le plus personnalisé possible pour obtenir l'assurance dont ils ont besoin pour se sentir protégés. En tant que fournisseur leader en Suisse, notre objectif est d'utiliser les dernières technologies disponibles afin de tenir cette promesse envers nos clients », a déclaré **Thomas Baumann IT architect chez La Mobilière**. « Nous avons collaboré avec IBM pour appliquer les capacités d'IA sur leurs systèmes IBM z16 de confiance afin de traiter les recommandations d'offres d'assurance plus rapidement et avec davantage de précision. En débloquent des modèles de données cachés avec des fonctions d'IA basées sur le NLP en temps quasi réel tout en garantissant la confidentialité et la sécurité, nous avons constaté une précision de 94 % dans les résultats des prédictions. Ces résultats très prometteurs nous ont incités à intégrer cette technologie dans nos processus de souscription dans un avenir proche. »*

IBM Z prend en charge l'intégralité du cycle de vie de l'IA

[L'IA générative](#) fait l'objet d'un grand nombre d'innovations, notamment IBM Watsonx Code Assistant for Z [récemment annoncé](#). Il s'agit d'un nouveau produit reposant sur l'IA générative pour la modernisation des applications mainframe, qui permettra de traduire plus rapidement le COBOL en Java sur l'IBM Z et d'améliorer la productivité des développeurs sur la plateforme.

Cependant, pour de nombreuses entreprises, la première étape pour tirer de la valeur de l'IA aujourd'hui consiste à se concentrer sur l'ensemble du cycle de vie de l'IA, qui comprend également le réglage fin, l'inférence et le déploiement des modèles de machine learning et de deep learning. Pour que les entreprises tirent le meilleur parti de leurs investissements dans l'IA, nous pensons qu'elles doivent exploiter leurs données critiques. L'IBM z16 est conçu pour évaluer les transactions commerciales à grande échelle, offrant la capacité de traiter jusqu'à 300 milliards de requêtes d'inférence de deep learning par jour avec une latence de 1 ms^[1]. Cette échelle ouvre des opportunités significatives pour les cas d'usage de l'IA sur le mainframe, tels que la détection de la fraude, la lutte contre le blanchiment d'argent, la compensation et le règlement, les soins de santé et la modernisation des applications.

Pour en savoir plus sur la nouvelle suite d'offres en matière d'IA et commencer votre parcours dans ce domaine, vous pouvez consulter le [site Web AI on IBM Z](#) et lire [l'article d'analyste récemment publié par Karl Freund](#).

[En savoir plus sur l'IA sur l'IBM Z](#)

Les plans, orientations et intentions d'IBM peuvent être modifiés ou retirés à tout moment, à la discrétion d'IBM et sans préavis. Les informations sur les potentiels futurs produits et améliorations sont fournies pour donner une idée générale des buts et objectifs d'IBM et ne doivent pas être utilisées pour prendre une décision d'achat. IBM n'est pas tenue de fournir du matériel, du code ou des fonctionnalités basés sur ces informations.

Contacts Presse :

Weber Shandwick pour IBM

IBM

Gaëlle Dussutour

Tél. : + 33 (0)6 74 98 26 92

dusga@fr.ibm.com

Louise Weber

Tél. : + 33 (0)6 89 59 12 54

ibmfrance@webershandwick.com

[\[1\]](#) Disclaimer: Performance result is extrapolated from IBM internal tests running local inference operations in a z16 LPAR with 48 IFLs and 128 GB memory on Ubuntu 20.04 (SMT mode) using a [synthetic credit card fraud detection model](#) exploiting the Integrated Accelerator for AI. The benchmark was running with 8 parallel threads each pinned to the first core of a different chip. The lscpu command was used to identify the core-chip topology. A batch size of 128 inference operations was used. Results may vary.
